



PySpark

直播：基于**Python Spark**的大数据分析

国内首部：PySpark框架、零基础上手、项目驱动、原理+实操

CONTENTS

01

主讲老师

轩宇：五年互联网大数据一线研发经验。。。

02

课程简介

近两年来，大数据分析师的岗位需求非常大。。。

03

课程信息

面向人群
课程定价
学习方式

04

课程大纲

PySpark环境篇
PySpark基础篇
PySpark进阶篇

PySpark项目篇
BigData实战篇

主讲老师

轩宇

五年互联网大数据**一线**研发经验，大数据技术**实战**派，先后担任过大数据开发工程师、开发经理、项目经理、**大数据架构师**，有丰富的**项目经验**。熟悉大数据整体生态链（调研、架构、开发、运维、应用），擅长利用大数据技术进行数据分析。近年来一直专注于**分布式计算**与机器学习相关领域，是国内**最早**的Spark研究与使用者。目前专注于Spark平台的大数据处理，任国内某知名在线教育机构大数据专业**首席**讲师，负责公司所有大数据开发与数据分析专业的产品研发与教学和企业培训工作。



课程简介

PART ONE

近两年来，大数据分析师的岗位需求非常大，70%的岗位技能需要掌握Spark作为大数据分析工具，其中以Python语言的易学性、快速开发的等特性很好地满足了大数据分析师的职业技能要求，所以国内外企业急需既会Hadoop/Spark大数据人员，也希望会Python的数据分析人员。



课程简介

PART TWO

本课程以案例驱动的方式讲解如何基于Python语言进行Spark Application编程，完成数据获取、处理、数据分析及可视化方面常用的数据分析方法与技巧，通过这些实际案例让学员轻松掌握使用PySpark分析来自不同领域的的数据，随课代码及案例代码均会提供给学员自行实操。



课程信息

面向人群

有一定Hadoop和Spark框架基础要提升的学习者
尚不会使用基于Python的Spark大数据分析师从业者
想使用Python进行开发的大数据工程师



课程信息

课程定价

原价：¥ ~~799~~元，优惠价 ¥ **399** 元



课程信息

学习方式

在线直播，共8次课，每次2小时
每周2次（周三、六，晚上20:30 - 22:30）
提供在线视频，课后反复学习



课程大纲

1. **PySpark**环境篇
2. **PySpark**基础篇
3. **PySpark**进阶篇
4. **PySpark**项目篇
5. **BigData**实战篇



PySpark环境篇

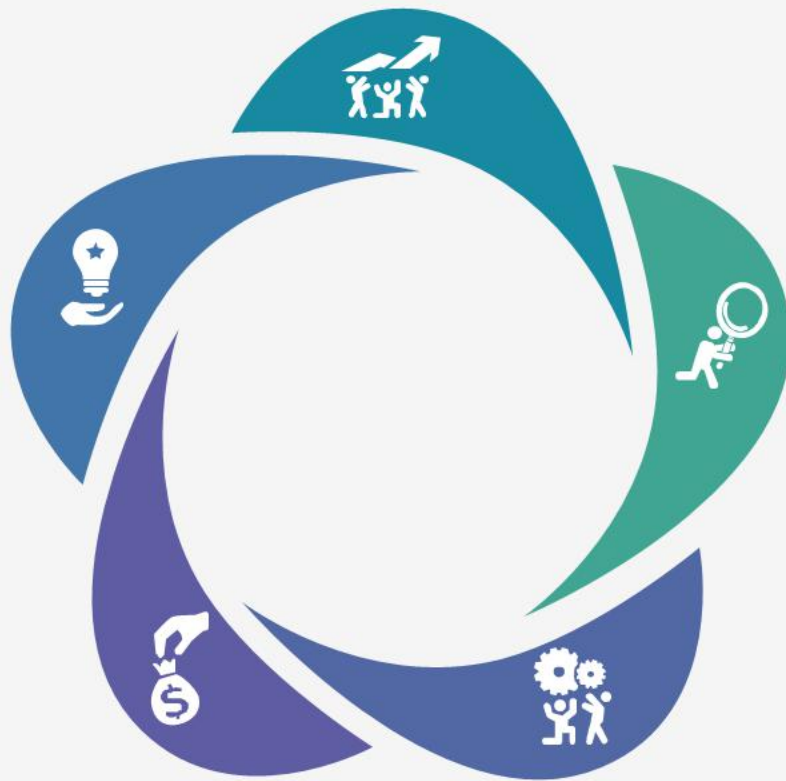
1、Python2.7+PyCharm环境配置

2、Anaconda环境配置

3、Python 数据结构及函数使用

4、Spark 2.x环境配置

5、Spark中pyspark命令行使用



PySpark基础篇

01
配置pyspark模块
到PyCharm

05
深入理解pyspark
内部工作原理



02
基于pyspark编程
实现WordCount
(词频统计)

03
spark-submit提交
Spark Application

04案例：
基于pyspark电商网站数据分析

PySpark进阶篇

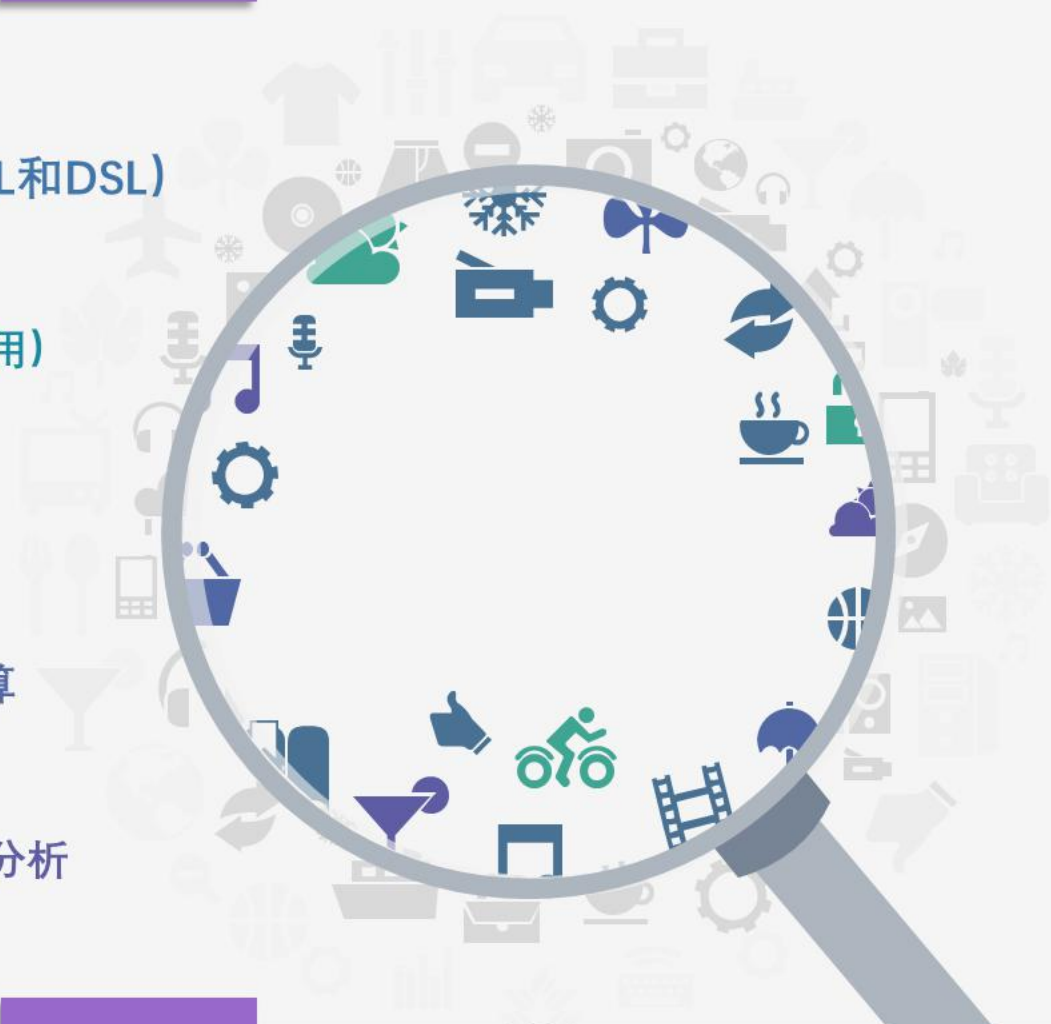
01 基于pyspark的SparkSQL编程（SQL和DSL）

02 案例：淘宝用户数据分析（集成Hive使用）

03 如何自定义、注册和使用UDF

04 基于pyspark的SparkStreaming实时计算

05 案例：实时销售额统计和商品点击趋势分析



PySpark项目篇

实战项目：
美国宇航局肯尼迪航天
中心WEB日志

数据统计

- 数据总览
- HTTP响应状态相关统计分析
- 客户端相关访问统计分析
- URI访问相关统计分析



数据调研

数据ETL

BigData实战篇



WELCOME LEARNING

THANKS